Supporting Information

The Regulation Mechanism for the Binding Between the SARS-CoV-2 Spike Protein and Host Angiotensin-Converting Enzyme II

Haiyi Chen[1,2], Yu Kang[2], Mojie Duan[1,*], Tingjun Hou[2,3*]

[1]National Centre for Magnetic Resonance in Wuhan, State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430071, Hubei, China

[2]Hangzhou Institute of Innovative Medicine, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China

[3]State Key Lab of CAD&CG, Zhejiang University, Hangzhou 310058, Zhejiang, China

*Corresponding address: mjduan@wipm.ac.cn; houtingjun@zju.edu.cn

Computational Methods

The crystal structure of the SARS-CoV-2 S-pro RBD in complex with ACE2 (PDB ID: 6LZG)[1] was utilized as the initial structure for the MD simulations. The Protein Preparation Wizard in Schrödinger 2017[2] was used to remove the crystallographic waters, ions, NAG, as well as add hydrogens and fix bond orders. The structure of DP4 was built by GaussView and processed by using the LigPrep module in Schrödinger 2017[2]. The RBD structure was extracted from the complex structure of S-pro and ACE2 to study the binding sites of DP4 on S-pro. The system containing the DP4-bound RBD and the interface of ACE2 was prepared for the enhanced sampling. Docking software *Glide*[3] was used to generate the complex of RBD-DP4. The binding box with the size of 20 × 20 × 20 Å centered on R355 was generated by using the *Receptor Grid Generation* component of *Glide*, and the fragment DP4 was docked to RBD using the standard precision (SP) scoring function of *Glide*.

The ff14SB force field[4] and general Amber force field (GAFF)[3] were used for the protein and DP4, respectively. The partial charges of the DP4 atoms were calculated by using the *antechamber* program in AMBER16 with the AM1-BCC method.[5] The TIP3P water molecules were added and the solute atoms were at least 15 Å away from the boundary of the rectangular box. The counter ions (i.e., sodium atoms and chlorine atoms) were added to neutralize the net charge of the system. Totally 100 replicas were prepared for the subsequent minimization, equilibration, and production runs. DP4 was randomly placed with the minimal distance from the atoms of S-pro larger than 10 Å in each replica. Each system was minimized, heated, equilibrated and simulated by the *pmemd* program in AMBER16[5]. First, a restraint force of 10.0 $kcal \cdot mol^{-1} \cdot Å^{-1}$ was exerted on the protein and ligand. The solvent and ion molecules were optimized by 5000 cycles of steepest descent and 5000 cycles of conjugate gradient minimizations. Then, the restraint was removed and the whole system was optimized by 5000 cycles of steepest descent and 5000 cycles of conjugate gradient minimizations. Next, each minimized system was heated from 10 to 300 K over a period of 25 ps, and then equilibrated over 0.5 ns in the NPT ensemble (T = 300 K and P = 1 bar). 100 ns of production run was conducted in the NPT ensemble. The temperature was controlled by the Langevin temperature equilibration scheme with a collision frequency of 1.0 ps^{-1}. The particle mesh Ewald (PME)[6] algorithm was used to handle the long-range electrostatic interactions under the periodic boundary condition, and a cutoff of 9.0 Å was used for the real-

space interactions. The SHAKE algorithm was used to constrain all covalent bonds involving hydrogen atoms, and the time step was set to 2 fs[7]. The final position of DP4 in each replica was collected using an in-house script.

Enhanced sampling was adopted to gain more insights about the thermodynamic profile of the binding process between the S-pro RBD and ACE2. In order to improve the sampling efficiency, only the interface region, including the residues 19-107 and 300-388 in ACE2 were extracted from the crystal structure, with a position restraint exerted on all the C_α atoms to maintain the basic conformation of the interaction interface of ACE2. Another system containing only RBD and the interface of ACE2 was also prepared. The same force fields and minimization protocols as those mentioned in the unbiased MD simulations were adopted for these two systems.

The metadynamics simulations were carried out by GROMACS 2018 with PLUMED2.5 plugin [8-10]. Two collective variables (CVs) were set to determine the sampling space. CV1 was set as the distance between the center-of-mass of heavy atoms in in the interface of RBD (residues K417, Y449, L455, F456, Y473, A475, E484, F486, N487, Y489, F490, Q493, Q498, T500, N501, G502, and Y505) and atoms in ACE2 (residue D30, D38, Q40, H34, K31, Q27, S19, Q24, L79, M82, Y83, F28, E35, Y41, Q42, L45, N330, D355, R357, K353, and G354). The COM distances as a function of simulation time were given in Figure S6. The CV2 was set as the contact value between residues in RBD and ACE2, which involve the residue pairs K417-D30, Y449-D38, Y449-Q42, Y449-H34, L455-D30, L455-K31, L455-H34, F456-Q27, F456-D30, F456-K31, Y473-Q27, A475-S19, A475-Q24, A475-Q27, E484-K31, F486-L79, F486-M82, F486-Y83, N487-Q24, N487-Y83, Y489-Q27, Y489-F208, Y489-K31, F490-K31, Q493-K31, Q493-H34, Q493-E35, Q498-D38, Q498-Y41, Q498-Q42, Q498-L45, T500-Y41, T500-L45, T500-N330, T500-D355, T500-R357, N501-Y41, N501-K353, G502-K353, G502-G354, G502-D355, Y505-E37, Y505-K353, and Y505-G354. The contact value for a conformation $\boldsymbol{X}$ sampled in the simulation is calculated as following[11-12]:

$$S(\boldsymbol{X}) = \sum_{(i,j)\in(i,j)_0} \frac{1}{1+\exp(\beta(r_{ij} - \lambda \cdot r_{ij}^0)} \qquad (1)$$

where β and λ were set to 50.0 nm^{-1} and 1.8, respectively, r_{ij} is the center-of-mass distance between residues *i* and *j*, r_{ij}^0 was set to 5.5 Å. The value S of a given conformation $\boldsymbol{X}$ sum over all the residue pairs *(i, j)* which are contact with each other in the native structure. The Gaussian width of CV1 and CV2 were set to 0.25 and 1.0, respectively. The Gaussian height was set to 0.5 kJ/mol. The bias factor, a parameter that decreases the added bias in a history dependent manner, was set to 32. Each system was sampled for 400 ns, and the added biases were deducted to recover the original free energy surface. The

convergency of the simulation was tested based on the free energy landscapes differences as a function of simulation time (Figure S7).

The free energy surface (FES) of each system was depicted using PLUMED2.5 plugin in GROMACS 2018[8-10]. Remarkable minima and saddle points were identified as the states in the 2-dimensional FES of each system, and those conformations belonging to each state were collected by an in-house script. All major contacts between the S-pro RBD and ACE2 residues in each state were collected with a cutoff of 5.5 Å by an in-house script, and were further divided into hydrophobic contacts and polar contacts based on the types of involved residues. All the polar interactions, including hydrogen bonds and salt bridges between the S-pro RBD and ACE2, were analyzed by the *cpptraj* program in AMBER16[5].

Table 1. Detail atoms of major hydrogen bonds in Figure 5 (> 50%). The atoms with labels in italic are corresponding to the S-pro RBD atoms.

region	Acceptor	Donor-H	Donor	Fraction (%)	Avg Dist (Å)	Avg Ang (°)
CR1	*N487@OD1*	Y83@HH	Y83@OH	84.6%	2.77	162.3
	A475@O	S19@HG	S19@OG	61.5%	2.70	161.4
	A475@O	S19@H1	S19@N	3.9%	2.87	135.4
CR2	E35@OE1	*Q493@HE22*	Q35@NE2	30.8%	2.77	168.3
	E35@OE2	*Q493@HE22*	Q35@NE2	23.1%	2.86	165.3
	D30@OD2	*K417@HZ2*	K417@NZ	38.5%	2.75	160.7
	D30@OD2	*K417@HZ3*	K417@NZ	11.5%	2.72	163.6
	D30@OD1	*K417@HZ2*	K417@NZ	7.7%	2.72	148.5
	D30@OD1	*K417@HZ3*	K417@NZ	7.7%	2.74	139.9
	D30@OD1	*K417@HZ1*	K417@NZ	3.9%	2.83	147.0
CR3	Y41@OH	*T500@HG1*	*T500@OG1*	65.4%	2.80	161.6
	Y41@HH	*T500@HG1*	*T500@OG1*	23.1%	2.88	150.9
	Y41@HE2	*T500@HG1*	*T500@OG1*	3.9%	2.91	147.6
	K353@O	*G502@H*	*G502@N*	76.9%	2.89	160.1
	E37@OE1	*Y505@HH*	*Y505@OH*	50.0%	2.68	166.1
	E37@OE2	*Y505@HH*	*Y505@OH*	11.5%	2.66	156.2

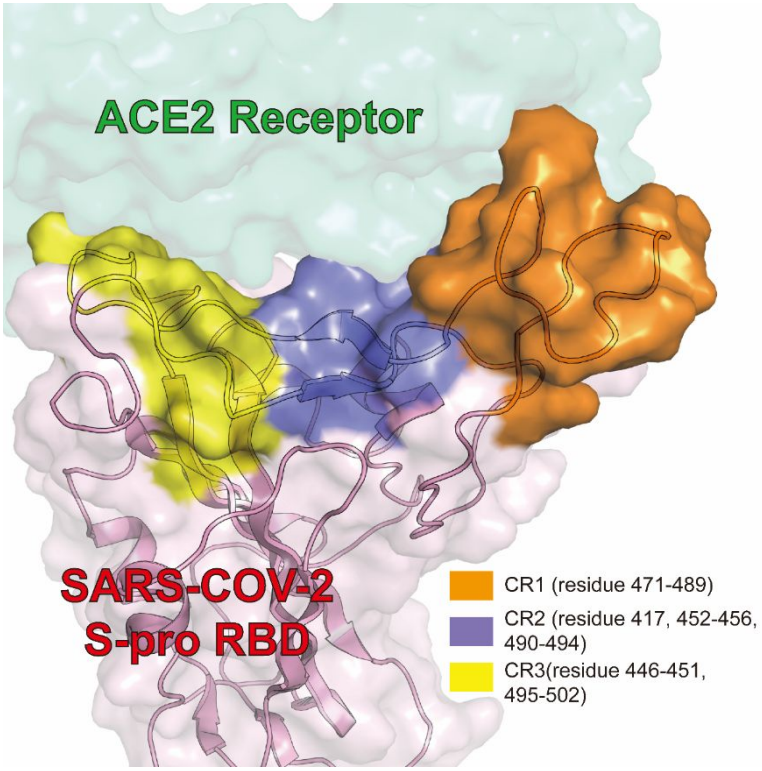


Figure S1. The different regions of binding interface on S-pro. The definition of regions on S-pro interface is according to the previous reported work[13].

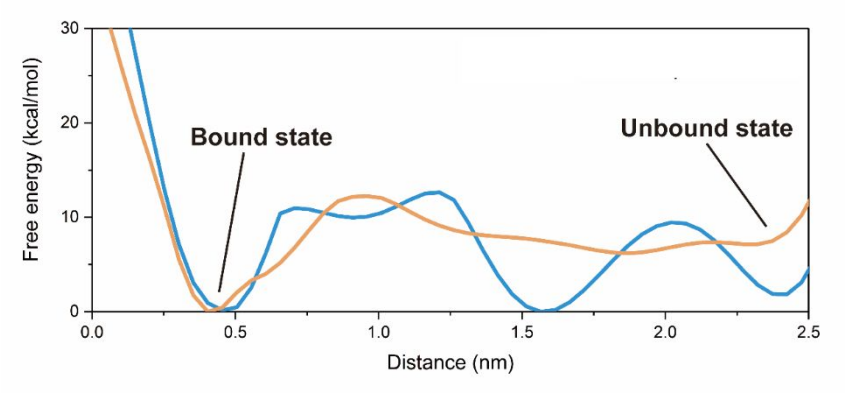


Figure S2. The one-dimensional free energy profile of binding process of apo-S-pro and ACE2 (blue line), and the free energy profile of binding process of S-pro/DP4 and ACE2 (brown line).

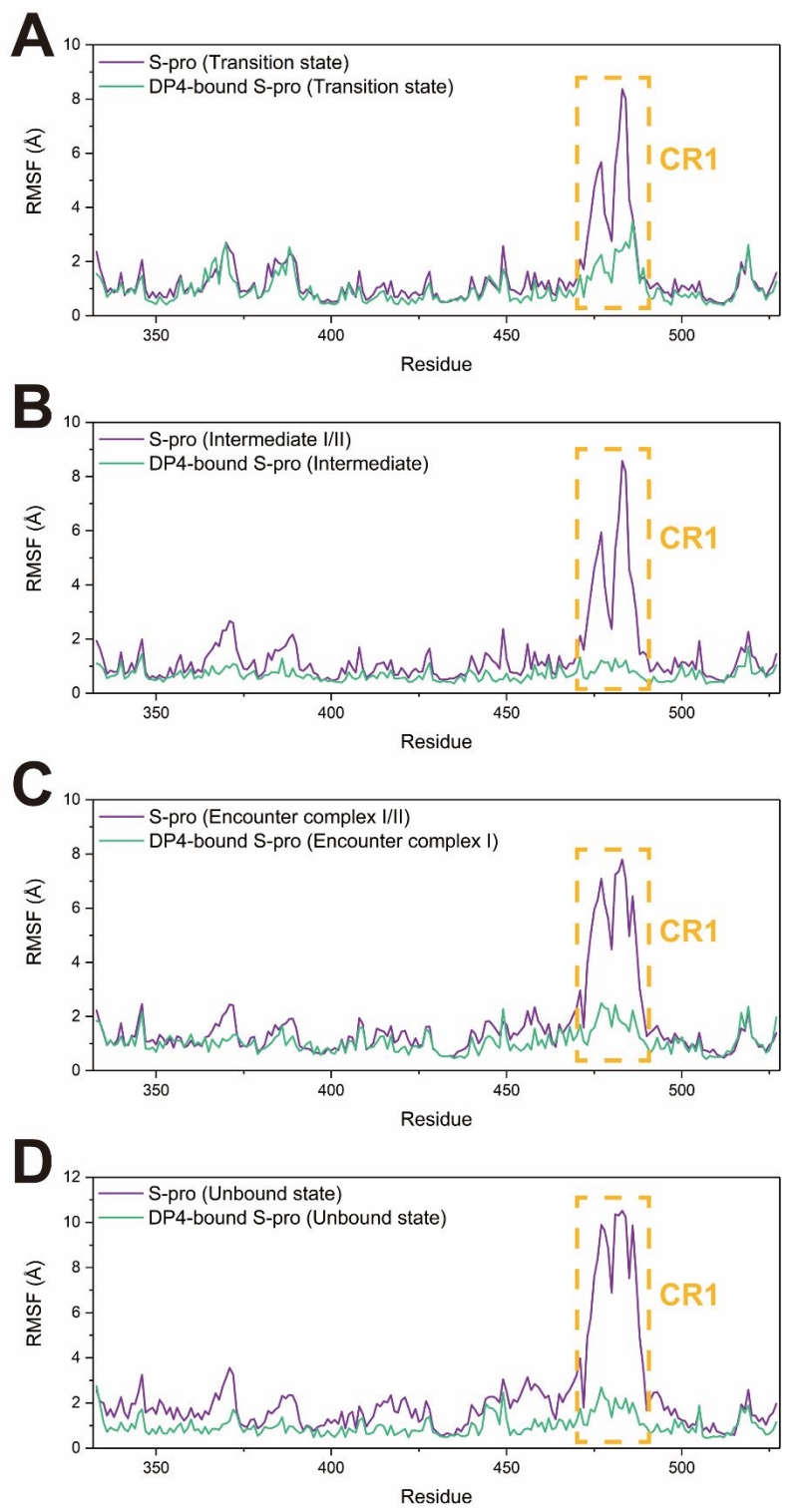
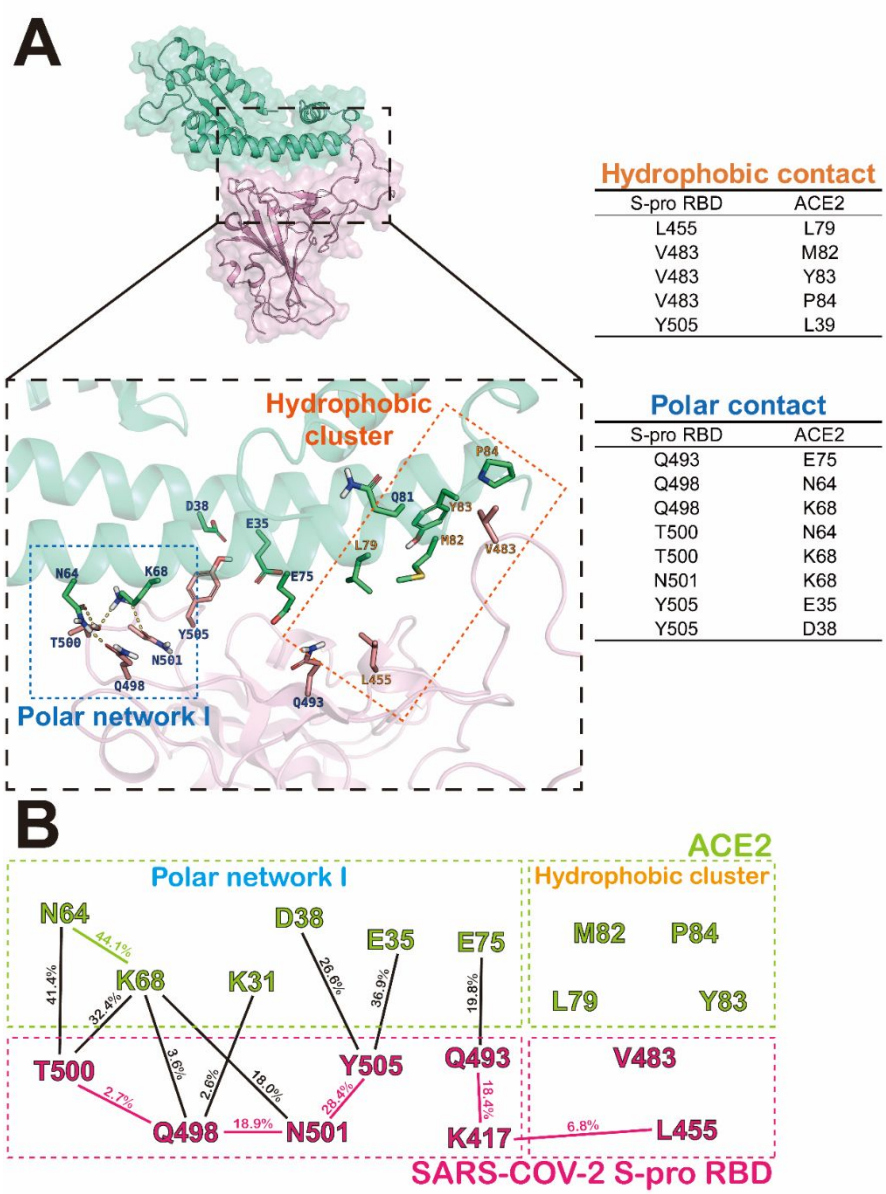


Figure S3. The structure fluctuation of S-pro in the binding process. (A) The heavy atom RMSF of structures in transition states. The CR1 regions (T470-F490 loop) are labeled by the yellow dashed frame. (B) The heavy atom RMSF of structures in intermediate states. (C) The heavy atom RMSF of structures in encounter complex. (D) The heavy atom RMSF of structures of DP4-bound/DP4-bound S-pro conformations in unbounded state.



Hydrophobic contact

S-pro RBD	ACE2
L455	L79
V483	M82
V483	Y83
V483	P84
Y505	L39

Polar contact

S-pro RBD	ACE2
Q493	E75
Q498	N64
Q498	K68
T500	N64
T500	K68
N501	K68
Y505	E35
Y505	D38

Figure S4. The interactions between S-pro and ACE2 in the encounter complex of system bounded with DP4. (A) The residues in the interaction interface of the conformation in encounter complex, the contact residue pairs with occupancy larger than 90% are listed in the figure. The interaction networks in different regions are labeled. (B) The interaction networks between RBD and ACE2. The residues on S-pro are colored in red, while residues in ACE2 are colored in green. The percentages of hydrogen bond formation are labeled on the lines between two residues.

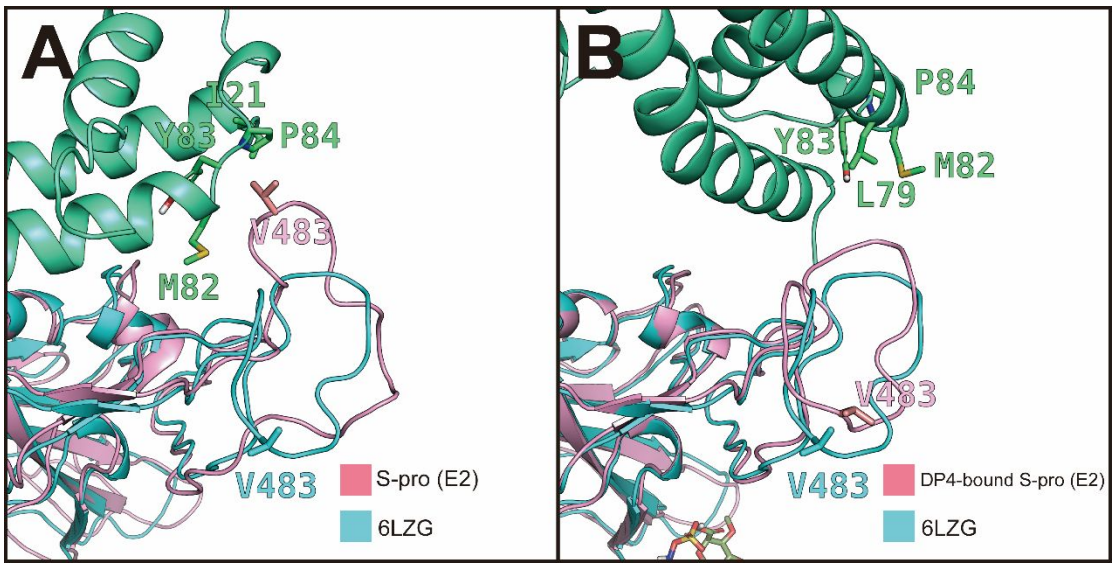


Figure S5. The interactions between T470-F490 loop of S-pro and ACE2 in the encounter complex of binding processes. (A) The representative structure and interactions in encounter complex of binding between apo-S-pro and ACE2. (B) The representative structure and interactions in encounter complex of binding between S-pro/DP4 and ACE2.

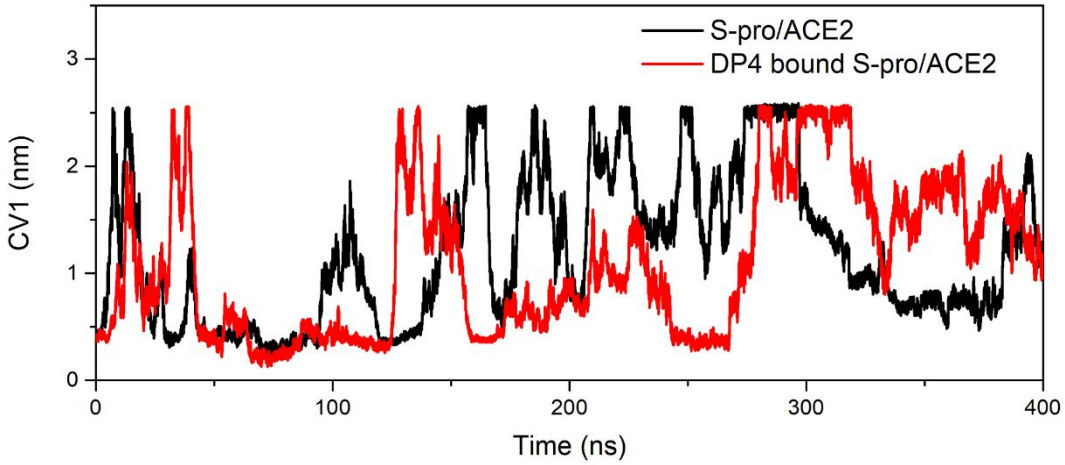


Figure S6. The center-of-mass distances (CV1) between S-pro and ACE2 as a function of simulation time. Several times of departing and re-binding of S-pro and ACE2 were observed in our simulations.

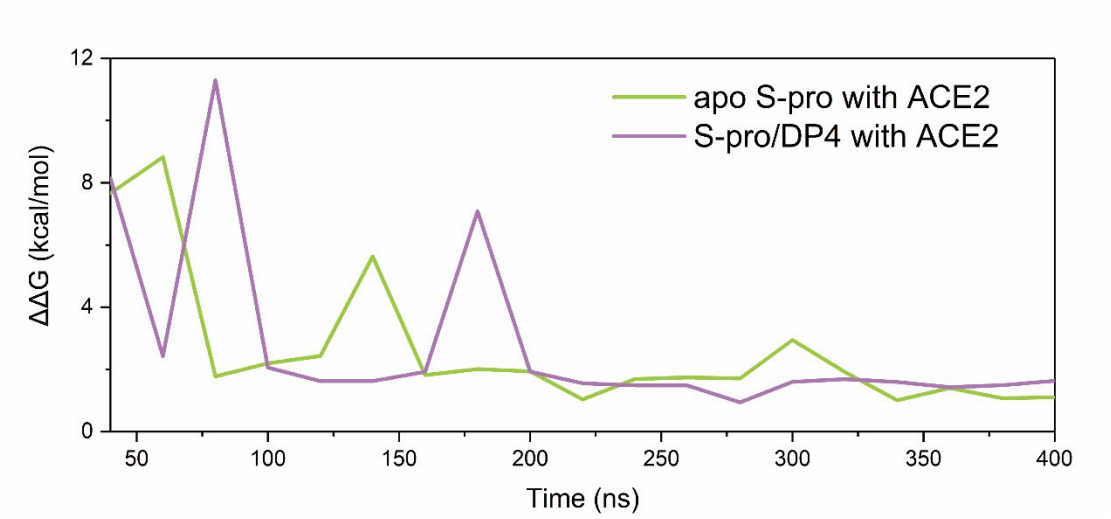


Figure S7. Convergence tests of the metadynamics simulations. The convergence was estimated by the free energy profile changes as a function of simulation time. ΔΔG is corresponding to the root-mean-square deviation of all states with low free energy (smaller than 100 kJ/mol) on the FES from time t-δt to *t*, where δt=20 ns. The ΔΔG values are lower than 2 kcal/mol after 200 ns indicate the convergence of the simulations.

References

1. Wang, Q. H.; Zhang, Y. F.; Wu, L. L.; Niu, S.; Song, C. L.; Zhang, Z. Y.; Lu, G. W.; Qiao, C. P.; Hu, Y.; Yuen, K. Y.; Wang, Q. S.; Zhou, H.; Yan, J. H.; Qi, J. X. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **2020,** *181* (4), 894-904.

2. *(2017) SchrÖdinger, version 2017, SchrÖdinger, LLC, New York, NY;* https://www.schrodinger.com, 2017.

3. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004,** *25* (9), 1157-1174.

4. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015,** *11* (8), 3696-3713.

5. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2013,** *3* (2), 198-210.

6. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.

7. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *Journal of Computational Physics* **1977,** *23* (3), 327-341.

8. Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1*, 19-25.

9. Pall, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. *Solving Software Challenges for Exascale* **2015,** *8759*, 3-27.

10. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014,** *185* (2), 604-613.

11. Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110* (44), 17874-17879.

12. Liu, N.; Guo, Y.; Ning, S. B.; Duan, M. J. Phosphorylation regulates the binding of intrinsically disordered proteins via a flexible conformation selection mechanism. *Communications Chemistry* **2020,** *3* (1), 123.

13. Wang, Y. J.; Liu, M. Y.; Gao, J. L. Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and interactions. *Proceedings of the National Academy of Sciences of the United States of America* **2020,** *117* (25), 13967-13974.